Just the Facts:
Technical Differences Between Talend Data Integration and Talend Open Studio

September 2017

1. Project and Source Control Management

No more developing in silos. With Talend Data Integration, you start by defining projects, users and roles in the Talend Administration Center. Next with source code management, your team does more by storing and sharing business models, data integration jobs, metadata, and documentation in a centralized repository (GIT, SVN). When you save your job, your colleagues immediately see your work and changes, and can check in/out the job for further updates.
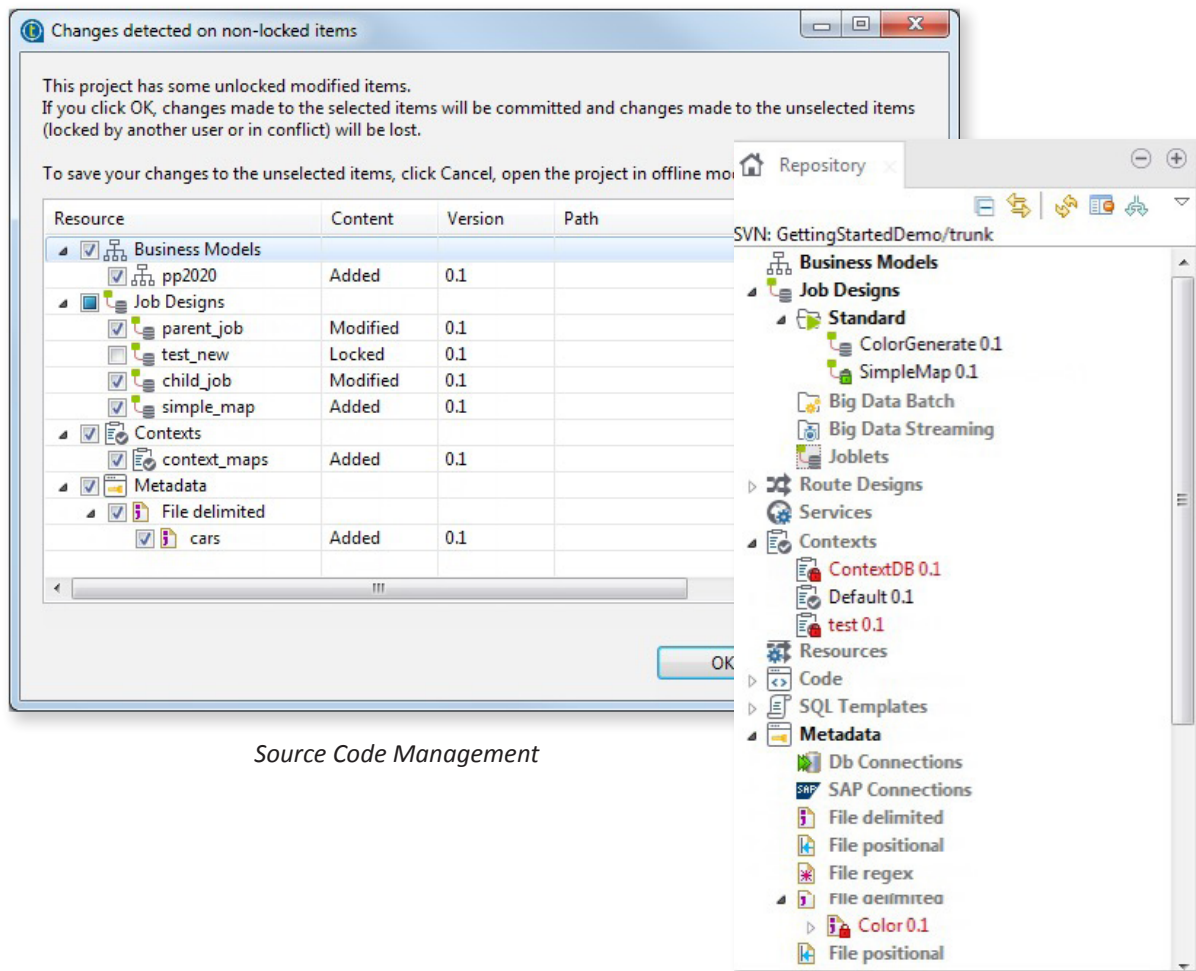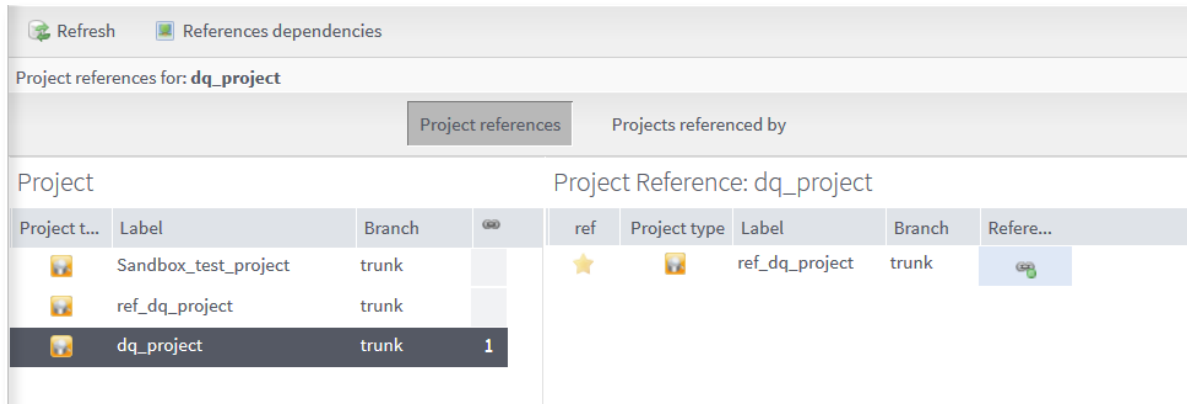

Source Code Management

Shared Repository View Using SVN

Reference Projects: To increase reuse and avoid duplication of items across jobs, reference projects of common objects and shared code can be created, which can then be included or linked to other projects. When the master changes, all slave projects are updated.
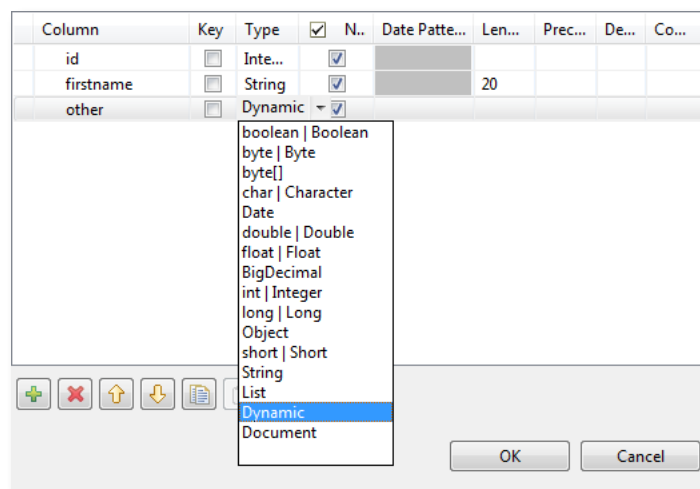
talend

Joblets: Simplify complex jobs by creating joblets which factorize job parts for reuse, like simple or complex transformations.
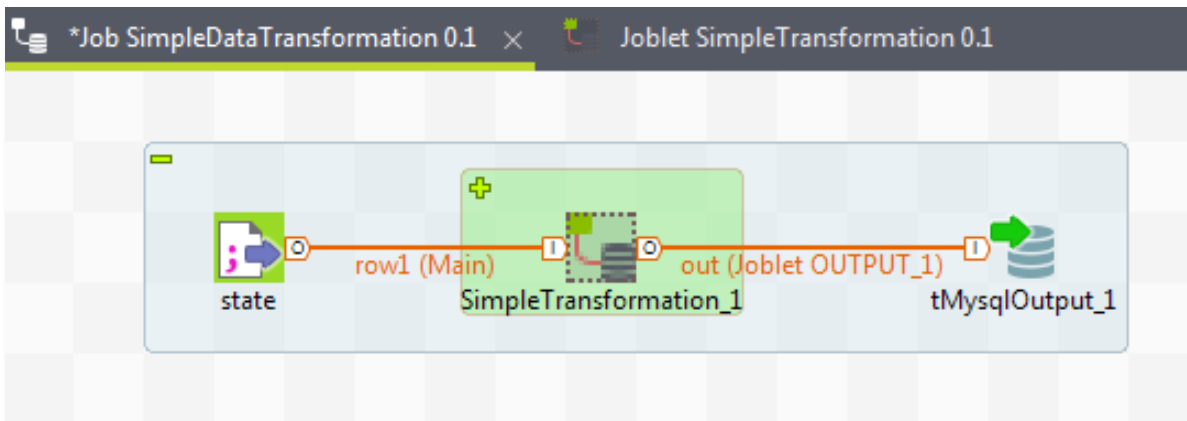

Dynamic Schemas: When the schema of the target is not known at design time, or when the target schema may vary depending on the specific instance of the database, dynamic schemas can be used to map source data to these targets "on-the-fly". This reduces the number of flows that need to be created and managed.
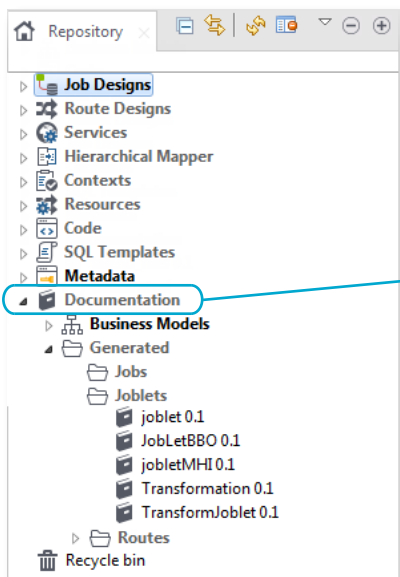
talend

Parallelization: To improve performance, you can partition the input data flow of a subjob into parallel processes on the same server to execute simultaneously.
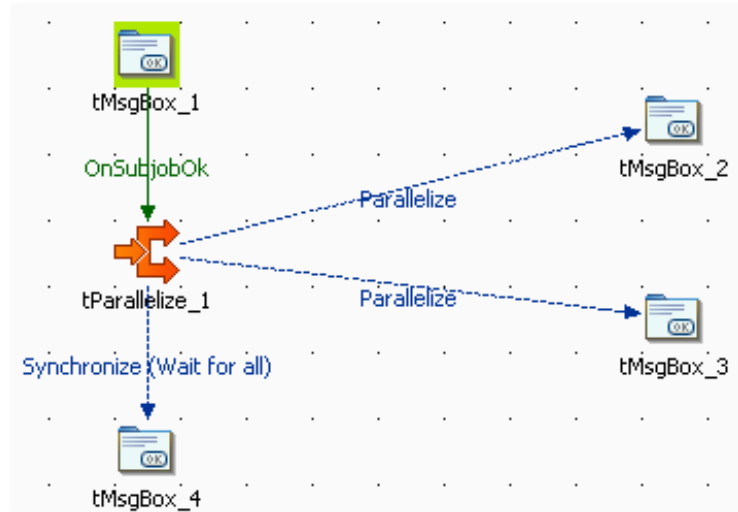
Documentation: Often a lower priority, but critical to supporting your project, you can have project documentation auto-generated each time the job is created, updated or saved. What a times savings!

Creating Test Cases: It is easy to create integration flow test cases, just click on the component and select "Create Test Case"
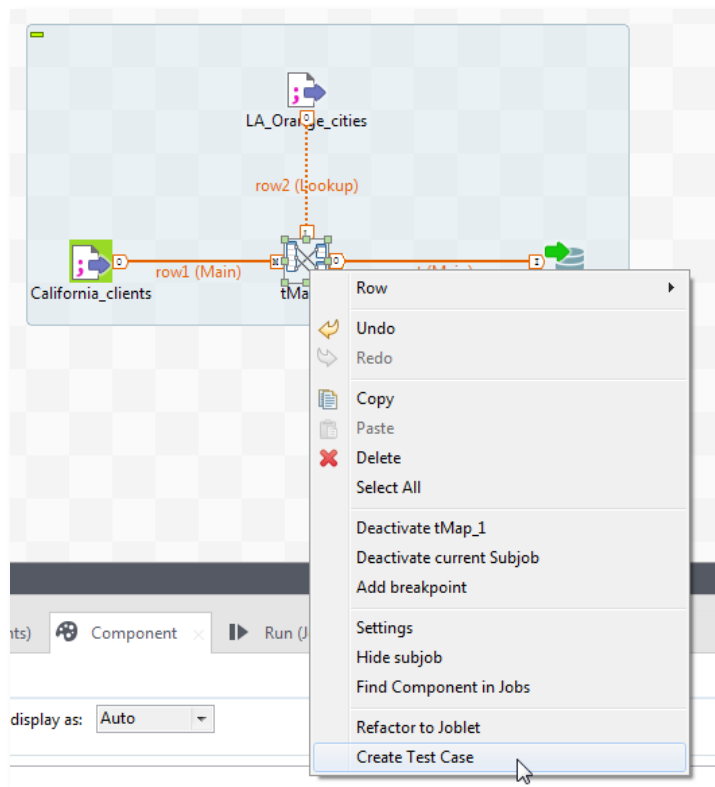
Continuous Integration: Get a significant DevOps productivity boost by automating build, test and deployment processes using Talend's built-in continuous integration (CI) features.

Team members integrate work frequently using Apache Maven-based builds that incorporate a distributed revision control system (GitHub), so everyone is working on the latest updates. Git support for graphical component-level diff and merge speeds updates, while automated tests (Junit) are generated to detect software defects earlier.

Higher quality code and fewer defects means faster time to production.

**Learn more about continuous integration** ❯
Video Documentation
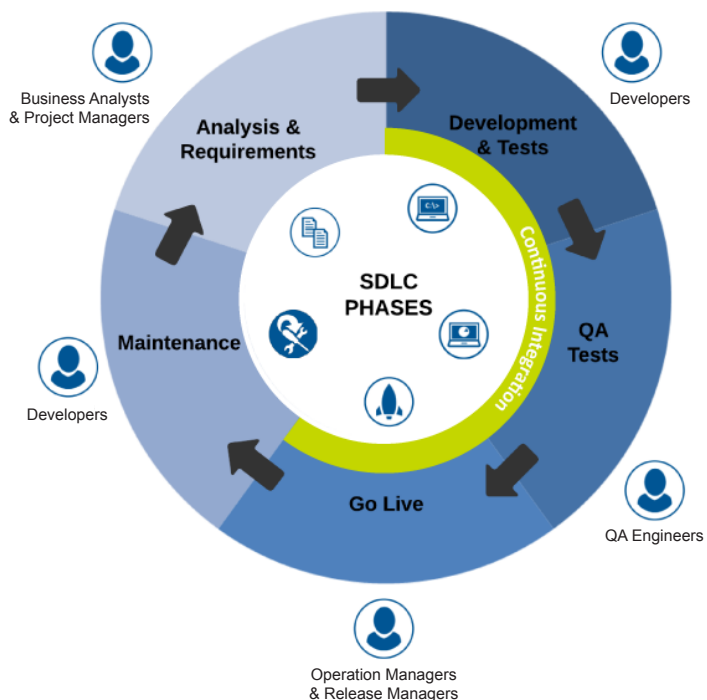
2. Job Monitoring and Event Logging

If your job is not performing well … do you know why?

With Talend's Activity Monitoring Console, event logging, and Kibana web-based job monitoring dashboards, you can see detailed information of all your processes, down to the component level, to help you quickly identify bottlenecks and optimize jobs for best performance. Debug the execution of integration actions, collect exception data and transfer it to the Studio console or web dashboard.

2.1 Activity Monitoring Console

The Talend Activity Monitoring Console monitors your end-to-end job flow including events (successes, failures, warnings, etc.), execution times and data volumes. Through detailed monitoring you better understand the underlying component and job interaction, and can then optimize performance. For example, you can visual how long a component took to execute today, and compare execution time over the past week. **Learn more** >
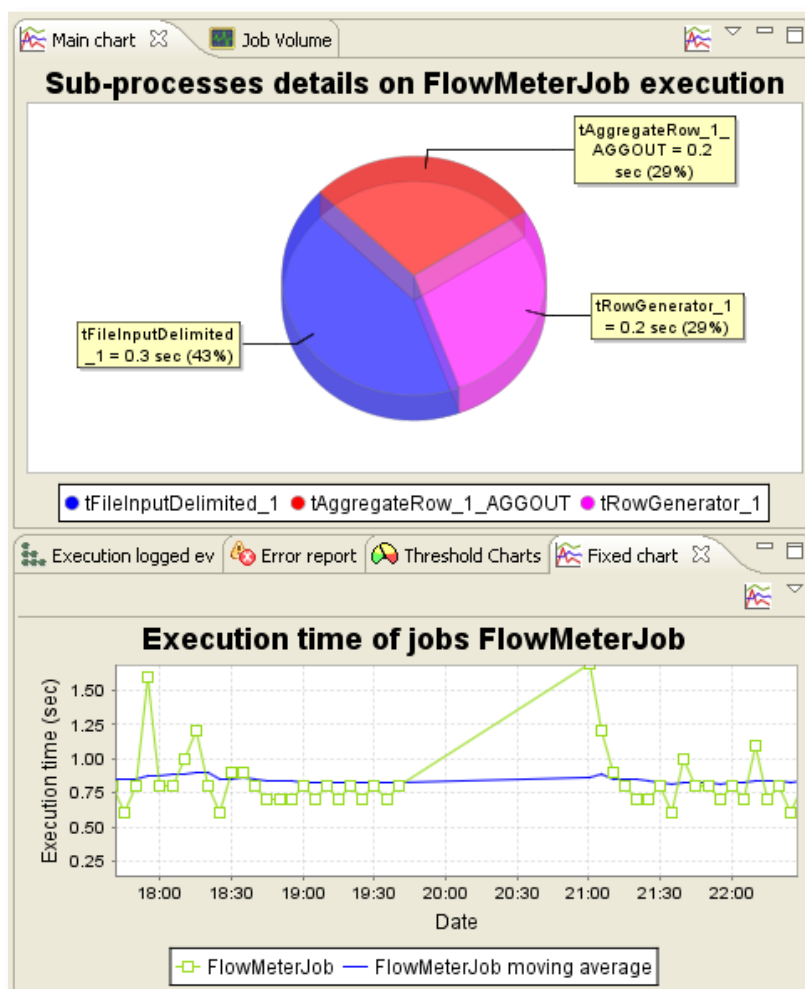

Analyze component performance in Activity Monitoring Console

2.2 Event Logging

The Activity Monitoring Console gives you the high-level process view, and Talend event logging tools give you the details. Rather than reading thousands of lines in a log file, integrated event logging based on ELK (Elasticsearch, Logstash and Kibana) provides powerful time-series data visualization with support for custom processing (custom filtering, data enrichment, aggregation). And with Apache Log4J services for each Talend component, you can get further enhanced diagnostics and error reporting on your dashboard.
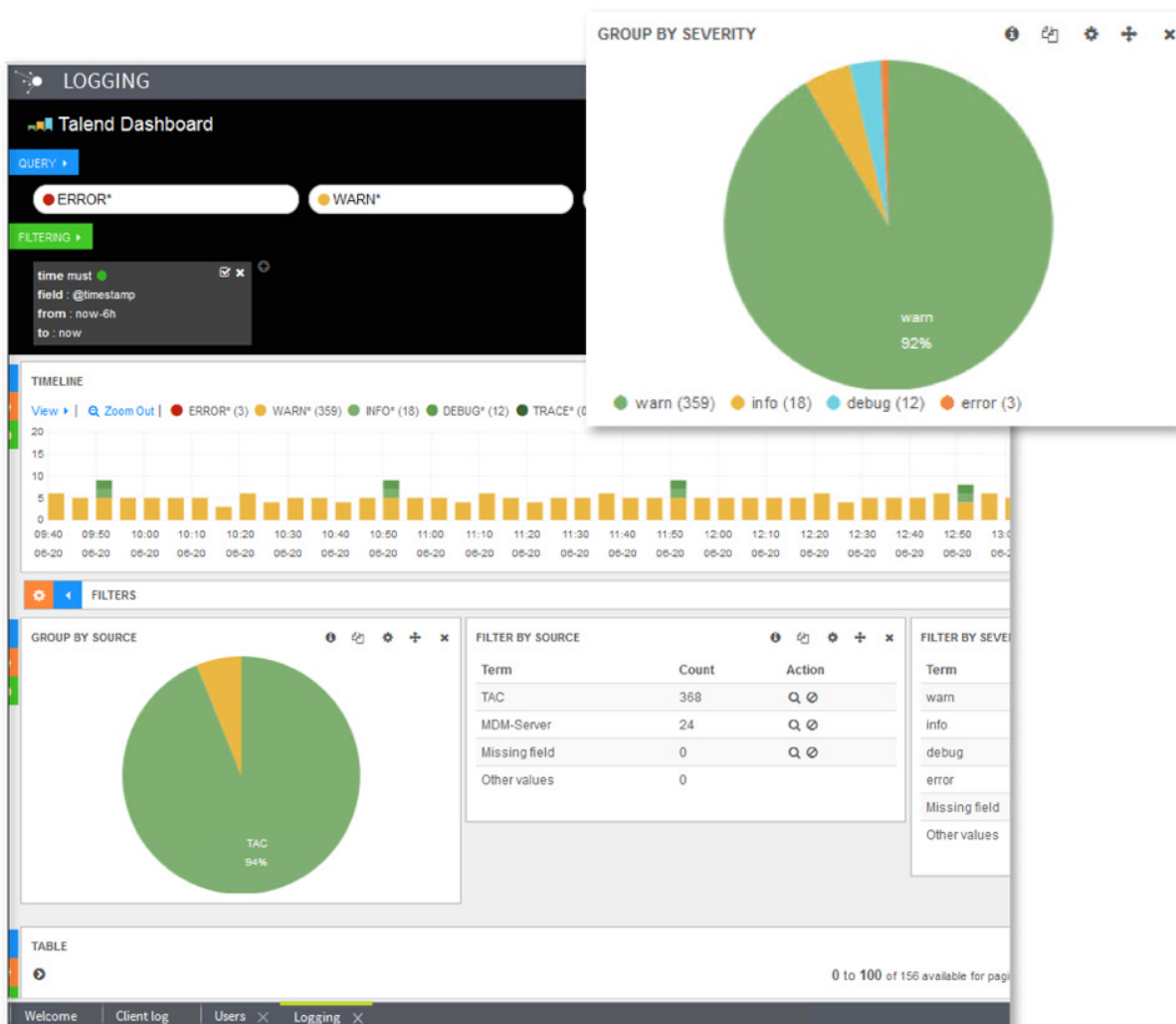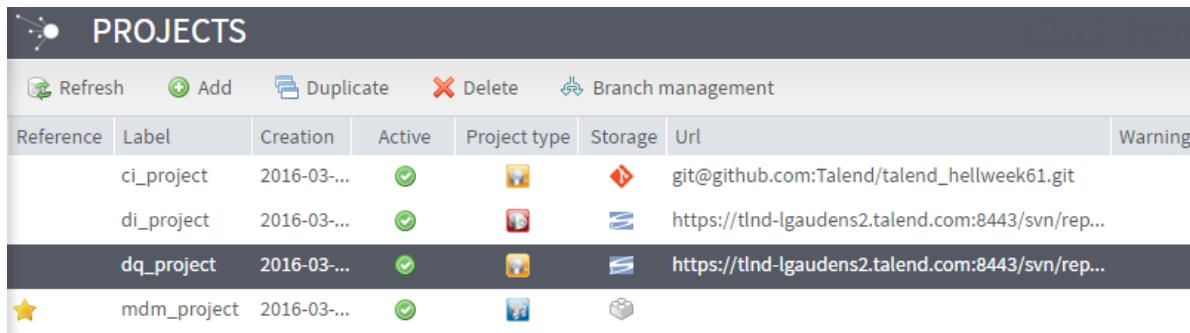

Use the default dashboards, or create customizable charts and graphs (e.g. pie chart, histogram, table) for a complete view of overall platform health, and then set up alerts so you can address issues before they happen. For example, you can build custom dashboards to monitor a unique product ID or database time-outs. **Learn more** >
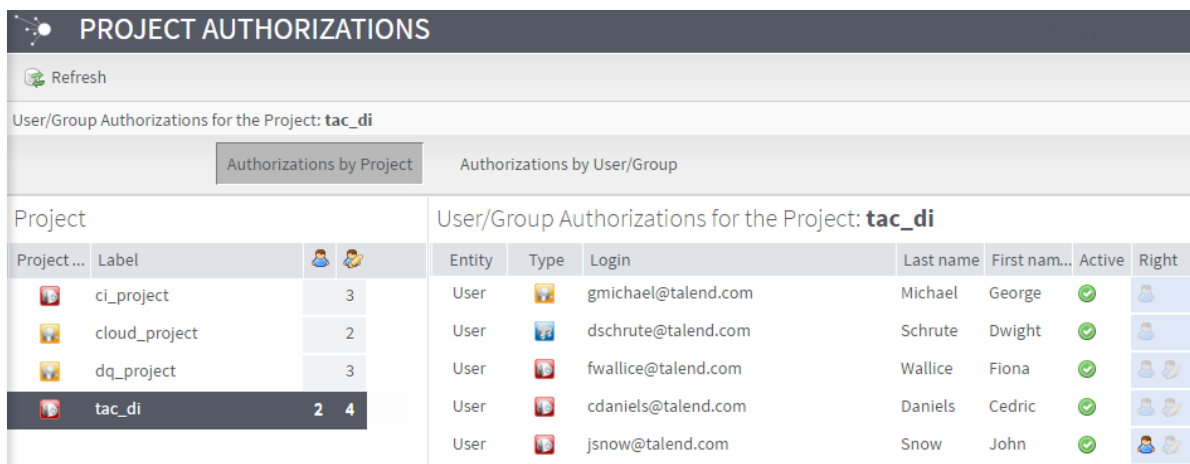
3. User and Deployment Management

Save significant time managing users and deployments with the Talend Administration Center, a web application where you can easily administer projects, users, access rights and licenses.
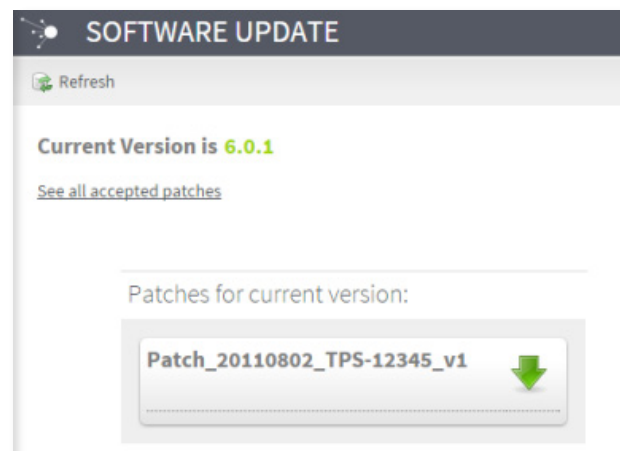

Set authorizations on users, projects and features - you have complete control of what team members can and cannot do.

Checking for and getting software updates is simplified, as it is done straight through the tool.
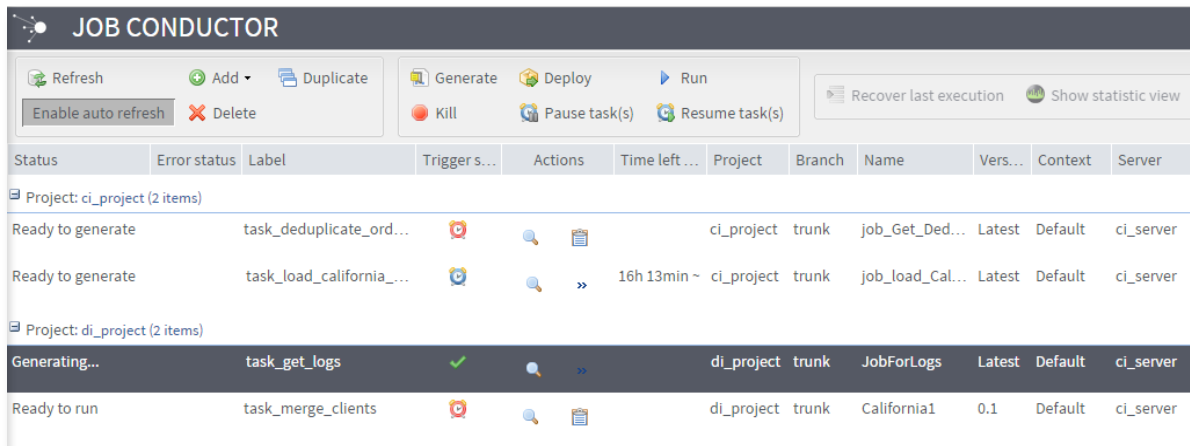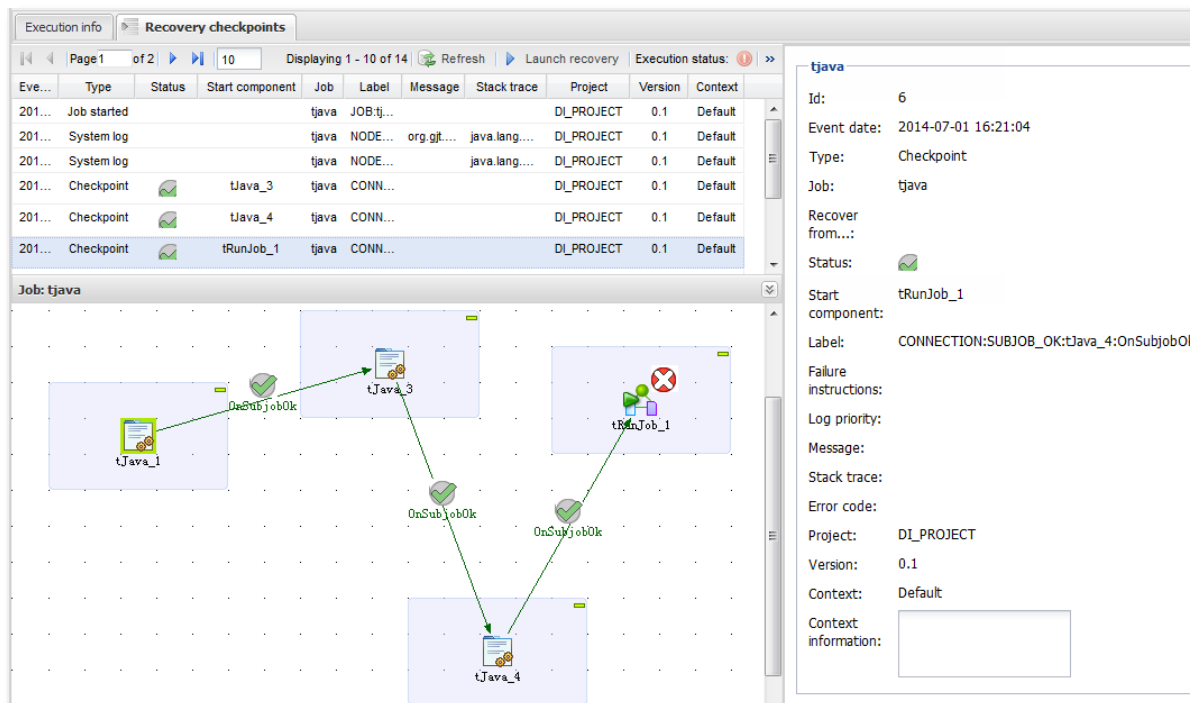
Job Scheduling: For deployment management, execution (job) servers are configured and managed in the graphical console. Next, graphically schedule and manage integration jobs with the Job Conductor, where you can define servers to execute on, time to start and stop, triggers for task execution order, resource (memory, process allocation) management, and more. Job scheduling can be parameterized with context variables or use the API for programmatic control.

Manage integration deployments with Job Conductor

Job Restart: If a job should fail, you can easily restart execution from the latest checkpoint before the failure.

Learn more about Talend Administration Center > *Video Documentation*

4. Integrated Data Preparation and Data Stewardship

Developers can save hours by allowing business users to access data and do basic cleansing and enrichment. End users know their data best, so let them do the work through an easy-to-use web interface, while IT governs the use of data and facilitates collaboration across the enterprise.

Talend Data Preparation enables anyone to access and cleanse data using browser-based, point-and-click tools. Smart guides and visual tools help anyone quickly understand data attributes and quality status. Results can easily be exported into desktop tools such as Tableau and Excel, enabling immediate decision-making. With one click you can then embed data preparation recipes into data integration pipelines.

Talend Data Stewardship provides robust data curation tools to help you quickly identify, manage, and resolve data integrity issues. Teams work together to establish a single version of the truth.

Learn more ❯ *Video Documentation*

5. And More ….

The appendix on the following page highlights all the differences between Talend Open Studio and Talend Data Integration. Depending on your requirements, there may be other Talend products that fit your needs, and they are all leverage the same Talend Studio you are familiar with.

- **If data quality is important, then Talend Data Management Platform extends Talend Data Integration with data quality and governance features.**

- **If you want to reduce deployment and management costs further, Talend Integration Cloud is a secure, managed integration Platform-as-a-Service (iPaaS), where you design jobs in Talend Studio, but instead of deploying and managing on-premises, you publish as a service to a secure AWS workspace managed by Talend.**

- **If you need to integrate big data sources and distributions and generate optimized Spark or MapReduce code, then check out Talend Big Data**

Appendix: A Comparison of Talend Data Integration Offerings

Talend 6.4

	Open Studio for Data Integration	Talend Data Integration	Talend Data Management Platform
License	Apache	Subscription	Subscription
Productivity Increasing Graphical Wizards and Tools			
900+ components and connectors	✓	✓	✓
Eclipse-based developer tooling	✓	✓	✓
ETL & ELT support	✓	✓	✓
Job designer	✓	✓	✓
Versioning	✓	✓	✓
Audit		✓	✓
Auto-generation documentation		✓	✓
Change data capture (CDC)		✓	✓
Drools business rule management system		✓	✓
Distant run		✓	✓
Dynamic schema		✓	✓
Interactive data viewer		✓	✓
Metadata bridge for metadata import/export		✓	✓
Parallelization		✓	✓
Re-usable joblets		✓	✓
Advanced SAP integration		✓	✓
Testing, debugging & tuning tools		✓	✓
Wizards		✓	✓
Visual mapping for complex XML and EDI			✓

	Open Studio for Data Integration	Talend Data Integration	Talend Data Management Platform
Improved Team Productivity			
Continuous delivery data integration		✓	✓
Impact analysis		✓	✓
Jobs compare		✓	✓
Reference projects		✓	✓
Team collaboration with shared repository		✓	✓
Centralized metadata management		✓	✓
Repository manager			✓
Monitor and Manage Complex Deployments with Ease			
Amazon EC2 lifecycle control		✓	✓
Check points, error recovery		✓	✓
Context management (dev, QA, prod)		✓	✓
Deployment manager		✓	✓
Execution plan, time and event-based scheduler		✓	✓
Log server with dashboard		✓	✓
Activity Monitoring Console		✓	✓
Talend Administration Center		✓	✓
High availability, load balancing, failover			✓
Accelerate Data Usage with Data Preparation and Stewardship			
Import, export & combine data from any database, Excel or CSV file		✓	✓
Import, export & combine CSV, Parquet & AVRO files from/ to HDFS		✓	✓

talend

	Open Studio for Data Integration	Talend Data Integration	Talend Data Management Platform
Export to Tableau		✓	✓
Self-service on-demand access to sanctioned datasets		✓	✓
Share data preparations & datasets		✓	✓
Operationalize preparations into any data & cloud integration flow		✓	✓
Run preparations on Apache Beam		✓	✓
Auto-discovery, profiling, smart suggestions, and data visualization		✓	✓
Auto-discovery & auto-profiling of custom semantic types		✓	✓
Smart & selective sampling & full-runs		✓	✓
Data tracking & masking with role-based security		✓	✓
Cleansing and enrichment functions		✓	✓
Data Stewardship App for data curation and certification		✓	✓
Data Quality and Governance			
Data profiling & analytics with graphical charts & drilldown data			✓
Automate data quality error resolution and enforce rules			✓
Data masking			✓
Data quality portal with monitoring, reporting & dashboards			✓
Semantic discovery with automatic detection of patterns			✓
Comprehensive survivorship			✓
Data sampling			✓
Enrichment, harmonization, fuzzy matching & de-duplication			✓

talend

	Open Studio for Data Integration	Talend Data Integration	Talend Data Management Platform
Advanced Data Profiling			
Fraud pattern detection using Benford Law			✅
Advanced statistics with indicator thresholds			✅
Column set analysis			✅
Advanced matching analysis			✅
Time column correlation analysis			✅
Reduce Risk with Product Indemnification		✅	✅

talend